24.5 An On-Chip Delay- and Skew-Insensitive Multi-Cycle Communication Scheme

Peter Caputa, Christer Svensson

Linköping University, Linköping, Sweden

Process shrinking of integrated circuits improves transistor performance while significantly increasing wire delays, now ranging up to multiple clock cycles. Repeater insertion along global interconnects is no longer sufficient; one is now forced to pipeline the wires by inserting flip-flops [1]. At the same time, global clock skew constraints, not only between blocks but also along pipelined interconnects, become even tighter. This development further complicates the design and verification process of high-performance VLSI chips [2].

A synchronous latency-insensitive design (SLID) method was proposed to manage the problems associated with global wire delays [3]. In the present paper, the implementation of this method is presented and its high performance and high robustness properties are demonstrated. The proposed technique not only mitigates unknown global wire delays, but also removes the need for controlling global clock skew. A successful implementation of a SLID-based 5.4mm-long on-chip global bus, supporting 3Gb/s/wire and accepting ±2 clock cycles of data-clock skew, in a standard 0.18μm CMOS process, is reported.

The SLID scheme is based on source-synchronous data transfer between blocks and data retiming at the receiving block. What is new about the technique is that all data is aligned to the correct receiver clock cycle, independent of clock skew and data delays (within limits). During the high-level phase of a SLID-design, the maximum delay+skew of a bus is estimated and a fixed n clock-cycles delay, covering the estimated delay+skew, is inserted in. After clock-cycle true verification, the n-cycle delay is implemented as a dual-port m-word FIFO-synchronizer (m>n), as shown in Fig. 24.5.1. This circuit aligns the incoming data to the local clock utilizing a single strobe signal (clock of the transmitting block) routed along the communication link. Incoming data is written into the FIFO at an address given by an input counter clocked by the strobe. Data is read from the FIFO at an address given by an output counter clocked by the local clock. By circular addressing and by maintaining an offset between the two counters, write and read never collide. The FIFO re-timer thus corresponds to standard wire pipelining through n clocked repeaters.

Overall global synchronization and clock alignment is achieved by relating each transmitted word to the strobe and local clock cycle with the same enumeration at each block, thus guaranteeing exactly n clock-cycles of latency for each link. This is accomplished by resetting the write and read counters to 0 and (m-n), respectively during a global asynchronous reset with no clock running. The input (output) counter thus counts up from 0 (m-n) as the first strobe (local clock) pulse reaches the write (read) port of the receiver at each block. Figure 24.5.2 shows how the input and output pointers, for the 4-word FIFO-retiming block in Fig. 24.5.1, progress during circuit operation. Note that the in-pointer time scale relates to the strobe (transmitter clock) and the out-pointer time scale relates to the local (receiver) clock.

A test chip is implemented to demonstrate the feasibility of the proposed concept. PRBS data is generated on-chip, for each bus wire, by transmitters (Tx-PRBS), which can also be set-up as circular shift registers. Data is sent at double the data-rate from a transmitting block to a receiving block across a 3b 5.4mm-long fully shielded repeaterless transmission-line-style global bus routed in metal6 (grounded metal4-plane as return path) and designed according to the principles in [4]. Phase-adjustable transmitter and receiver clocks are generated either on-chip, utilizing voltage-controlled ring oscillators, or are delivered from off chip. The retiming circuitry at the receiving block is a dual-data-rate version of the m=4-word dual-port FIFO memory, shown in Fig. 24.5.1. Figure 24.5.3 shows a block diagram of the implemented FIFO-synchronizer write-port, where the received data, transmitted on the rising and falling clock edge respectively, is fed to two separate read ports. The schematic of the double-edge-triggered flip-flop, utilized in the input counter, is shown in Fig. 24.5.4. An on-chip self-test block compares the received data with reference PRBS-data (Ref-PRBS) generated in the receiving block. Figure 24.5.5 shows simulated waveforms of the received data and reference data at the self-test block input, when n=2.

For circuit evaluation, the fabricated SLID-communication link is reset with a write/read addressing delay (n) of 1, 2, and 3 clock cycles, respectively. Measurements showed error-free functionality for each setup achieving 3Gb/s/wire at nominal V_{dd}=1.8V, 25˚C utilizing a 1.5GHz external clock source and worst-case bit-patterns for capacitive and inductive coupling, respectively. The maximum data-rate is not limited by interconnect crosstalk effects, but rather by the circuitry for on-chip data generation and clock distribution.

A benefit of the proposed scheme is that clock and link delays can vary dynamically during system operation, without risk of communication failure. Two separate (but phase-locked) external clock sources, running at 500MHz, are utilized to measure clock-skew tolerance. The link is setup with a delay of 1, 2, and 3 clock cycles, respectively, and a repetitive bit pattern is transmitted across the bus. For each case, the receiver clock (local clock) is dynamically phase shifted with respect to the transmitter clock (strobe) until bit errors are registered in the received data. Figure 24.5.6 shows the measured maximum positive (local clock arriving before strobe) and negative (local clock arriving after strobe) clock-skew tolerance. The arrows show the experimental error-free region of the external skew, measured to ±2 clock cycles, for various settings of n. Figure 24.5.7 shows a chip micrograph of the circuit fabricated in a 1.8V 0.18μm 6M CMOS process [5].

The proposed SLID-communication link behaves as a fully synchronous system with pipelined interconnects. A significant benefit of the SLID-technique is that the requirement of synchronous clock distribution along the buses and low clock skew between blocks is completely removed.

References:

[1] P. Cocchini, "A Methodology for Optimal Repeater Insertion in Pipelined Interconnects," *IEEE Trans. Computer-Aided Design*, vol. 22, pp. 1613-1624, 2003.

[2] R McInerney, et al., "Methodology for Repeater Insertion Management in the RTL, Layout, Floorplan and Fullchip Timing Databases of the Itanium Microprocessor," *Proc. Int'l. Symp. Physical Design*, pp. 99-104, 2000.

[3] A. Edman and C. Svensson, "Timing Closure through a Globally Synchronous, Timing Partitioned Design Methodology," *Proc of Design Automation Conference*, pp. 71-74, 2004.

[4] P. Caputa and C. Svensson, "Well-Behaved Global On-Chip Interconnect," *IEEE Trans. Circuits Syst. I*: vol. 52, no. 2, pp. 318-323, 2005.

[5] http://cmp.imag.fr/products/ic/?p=STHCMOS8, June, 2005

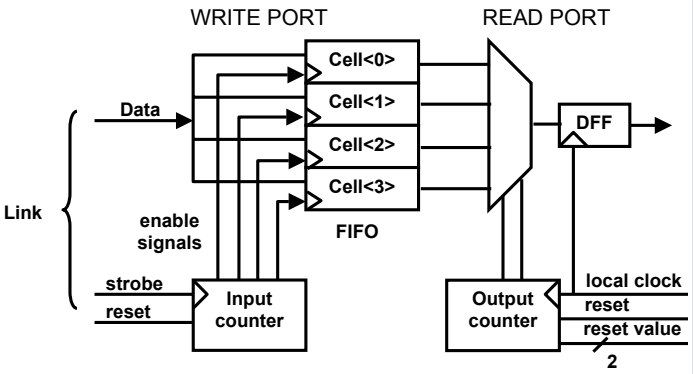

Figure 24.5.1: A dual-port 4-word FIFO synchronizer.

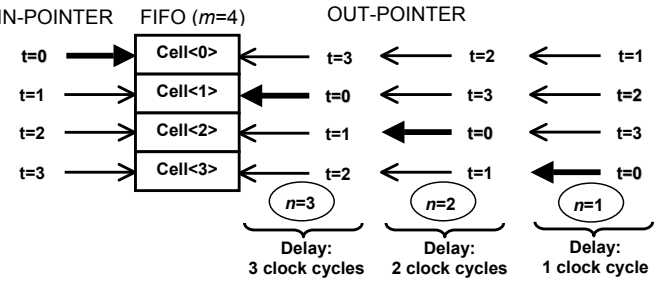

Figure 24.5.2: Progress of the FIFO in-pointer and out-pointer.

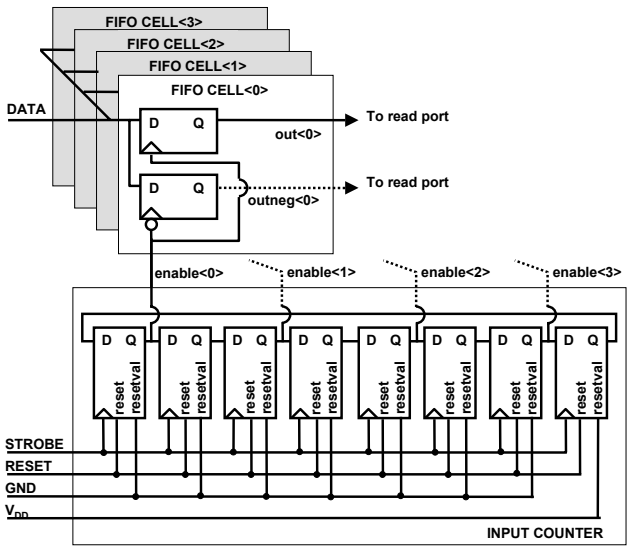

Figure 24.5.3: Block diagram of the FIFO-synchronizer write port.

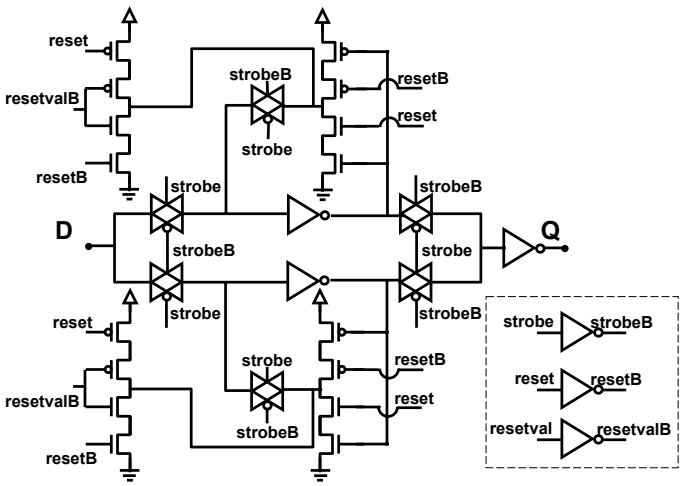

Figure 24.5.4: Schematic of the double-edge triggered flip-flop utilized in the input counter.

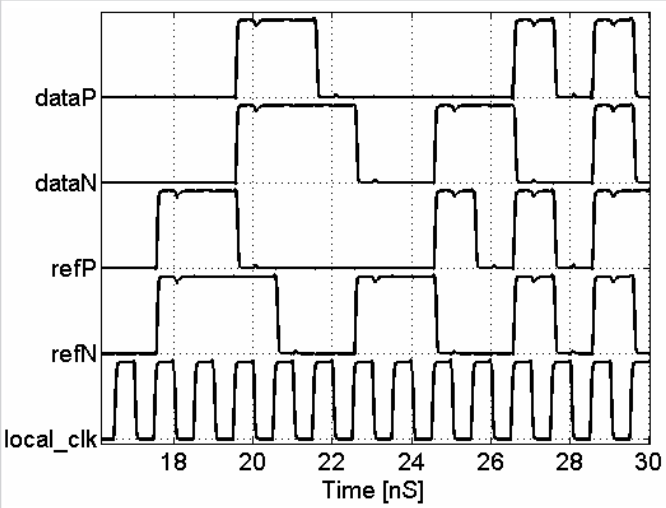

Figure 24.5.5: Simulated waveforms when n=2. dataP(dataN) is the received data transmitted on the rising (falling) clockedge while refP(refN) is the corresponding reference data.

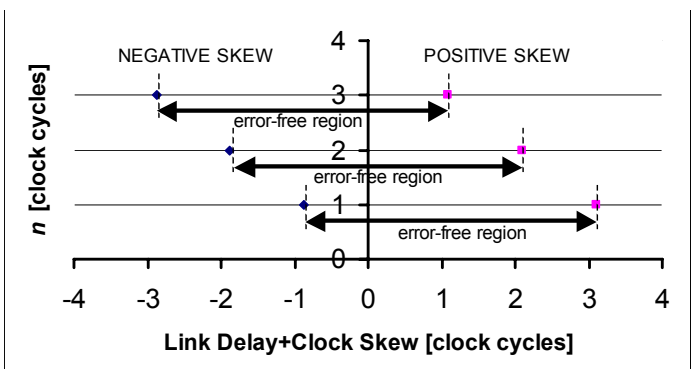

Figure 24.5.6: Measured maximum positive and negative link delay+clock skew tolerance.

Continued on Page 664

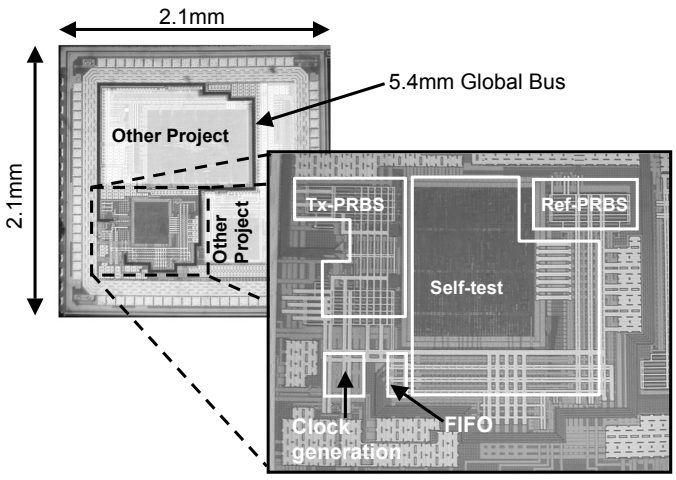

Figure 24.5.7: Micrograph of the test chip fabricated in 0.18µm CMOS.

